Combining genotype, phenotype, and environmental data to delineate site-adjusted provenance strategies for ecological restoration

Carolina S. Carvalho[1], Brenna R. Forester[2], Simone K. Mitre[1], Ronnie Alves[1], Vera L. Imperatriz-Fonseca[1], Silvio J. Ramos[1], Luciana C. Resende-Moreira[1], José O. Siqueira[1], Leonardo C. Trevelin[1], Cecilio F. Caldeira[1], Markus Gastauer[1], Rodolfo Jaffé[1,3*]

[1] Instituto Tecnológico Vale, Belém-PA, Brazil.

[2] Colorado State University, Fort Collins-CO, USA

[3] Departamento de Ecologia, Universidade de São Paulo, São Paulo-SP, Brazil.

[*] **Correspondence:**

Rodolfo Jaffé

email: r.jaffe@ib.usp.br

phone: +55 (91) 3213 5523

Abstract

Despite the importance of climate-adjusted provenancing to mitigate the effects of environmental change, climatic considerations alone are insufficient when restoring highly degraded sites. Here we propose a comprehensive landscape genomic approach to assist the restoration of moderately disturbed and highly degraded sites. To illustrate it we employ genomic datasets comprising thousands of single nucleotide polymorphisms from two plant species suitable for the restoration of iron-rich Amazonian Savannas. We first use a subset of neutral loci to assess genetic structure and determine the genetic neighborhood size. We then identify genotype-phenotype-environment associations, map adaptive genetic variation, and predict adaptive genotypes for restoration sites. Whereas local provenances were found optimal to restore a moderately disturbed site, a mixture of genotypes seemed the most promising strategy to recover a highly degraded mining site. We discuss how our results can help define site-adjusted provenancing strategies, and argue that our methods can be more broadly applied to assist other restoration initiatives.

Keywords: Genotype-environment associations (GEA), genotype-phenotype associations (GPA), landscape genomics, local adaptation, RAD sequencing, restoration genomics, single nucleotide polymorphisms (SNP).

Introduction

In spite of the broadly recognized importance of genetic provenance for restoration initiatives, the use of genomic tools to define provenance strategies is still uncommon. Choosing provenances based on genetic knowledge can help increase genetic diversity and adaptability, thereby contributing to the success of restoration initiatives (Broadhurst et al., 2008; Mijangos et al., 2015; Weeks et al., 2011). Fortunately, the use of genetic provenancing is increasing, as advances in Next-Generation-Sequencing technologies have made possible large-scale assessments of neutral and adaptive genetic variation (Breed et al., 2019; Mijangos et al., 2015; Williams et al., 2014). For instance, neutral loci (i.e., those which are not subject to natural selection) can be used to identify independent demographic units, assess fine-scale spatial genetic structure, and quantify genetic diversity (Allendorf et al., 2013; Balkenhol et al., 2017); whereas adaptive loci (under natural selection) are relevant to detect adaptations to local environmental conditions and delineate adaptive units (Funk et al., 2012; Rellstab et al., 2015).

Restoration genomic studies published so far have assessed the effect of multiple environmental variables on genetic composition in order to identify which individuals or populations are "pre-adapted" to future climates (Gugger et al., 2018; Lu et al., 2019; Martins et al., 2018; Rossetto et al., 2019; Shryock et al., 2017, 2015; Steane et al., 2014; Supple et al., 2018). Although this information is essential to inform predictive and climate-adjusted provenancing schemes (Prober et al., 2015), the emphasis on climate has overshadowed the application of genomic methods to restore extremely degraded sites (Bucharova et al., 2019; Lesica and Allendorf, 1999). Such site-adjusted provenancing (targeting the restoration of

specific sites considering their current environmental conditions) may not even incorporate climate change considerations, as highly degraded sites have unique characteristics that make them extremely challenging to restore. First, highly degraded sites usually require immediate restoration or rehabilitation, so adaptations to current environmental conditions are more suitable to guide provenance strategies than those based on future climate (Gastauer et al., 2019). Second, environmental protection agencies usually require the restitution of ecosystems to conditions as close to a pre-disturbance baseline as possible, as well as regular (and costly) monitoring until rehabilitation goals have been achieved (Gastauer et al., 2019). The main efforts thus lay in the quick establishment of viable populations that will restore ecosystem functions and processes, prevent soil erosion, and protect biological diversity. Third, highly degraded sites such as exhausted open-pit mines have radically different environmental characteristics than natural habitats (Gastauer et al., 2019), so site-specific characteristics are of primary importance to define provenance strategies. Such site-specific variables generally need to be measured *in situ* and at fine spatial resolution, since they may not be available as spatial layers in open-access repositories (and if they are, their spatial resolution may be too coarse to reflect the reality of environmental conditions on the ground). Finally, site-adjusted provenancing strategies need to consider local adaptations to climate, soil, terrain, and even biological interactions, whereas climate-centered provenancing strategies focus exclusively on climate.

The degree of disturbance can play an important role in determining site-adjusted provenancing strategies (Breed et al., 2013; Lesica and Allendorf, 1999). Whereas local genotypes are generally the best suited to restore sites where the degree of disturbance is low, adaptations found in distant populations may facilitate establishment in highly degraded sites

84 to which local genotypes may not be adapted (Breed et al., 2013; Broadhurst et al., 2008;
85 Lesica and Allendorf, 1999)). Mixtures of genotypes from different populations have been
86 suggested as the best strategy to recover highly degraded sites, given that enhanced genetic
87 variation is more likely to rapidly generate local adaptations to novel ecological challenges
88 (Lesica and Allendorf, 1999). In any case, determining the most appropriate site-adjusted
89 provenancing strategy will require the delineation of local provenances and the spatial
90 distribution of local adaptations (Breed et al., 2019).

91 Although the use of neutral genetic markers to identify independent demographic units
92 is now common practice (Coates et al., 2018), few restoration studies have delineated seed
93 sourcing strategies based on population genetic structure (Durka et al., 2017) or the genetic
94 neighborhood size (the distance at which genetic composition stops being spatially
95 autocorrelated) (Krauss et al., 2013; Krauss and Koch, 2004; Rossetto et al., 2019). On the
96 other hand, only three restoration genomic studies so far have identified putative adaptive loci
97 and then mapped adaptive genetic variation (Martins et al., 2018; Shryock et al., 2015; Steane
98 et al., 2014). While the assessment of phenotype through common garden and reciprocal
99 transplant experiments to identify local adaptations has a long history (Aitken and Bemmels,
100 2016), no study has yet combined genotype-environment associations (GEA) with genotype-
101 phenotype associations (GPA) to delineate seed sourcing areas. This approach could improve
102 the inference of potential candidate genes and provide important insights into genes
103 underlying fitness-related traits (Mahony et al., 2019; Talbot et al., 2016; Vangestel et al.,
104 2018).

105 Here we propose a comprehensive landscape genomic approach to assist the
106 restoration of moderately disturbed and highly degraded sites. Relying on genotyping-by-

sequencing we identified thousands of single nucleotide polymorphisms in two plant species of special interest for the restoration of exhausted mining sites from the Carajás Mineral Province, located in the Eastern Amazon (Skirycz et al., 2014; Souza-Filho et al., 2019). We first used a subset of neutral loci to assess broad and fine-scale genetic structure and determine the genetic neighborhood size. Subsequently, we combined univariate and multivariate methods to identify GEA and GPA and employed spatial principal components analyses (sPCA) to map adaptive genetic variation while accounting for spatial autocorrelation in genetic composition. Finally, we predicted the adaptive genotypes associated with the environmental conditions of restoration sites (Fig. 1).

We focus on the Carajás Mineral Province, which harbors one of the world's largest deposits of iron ore and huge iron ore mining projects, with operations dating back to the 1980s (Poveromo, 1999; Souza-Filho et al., 2019). The banded ironstone formations known as *Cangas* (where iron-ore deposits are concentrated) are characterized by shallow, acidic, nutrient-depleted and metal-rich soils and marked by high solar radiation, hot temperatures, and a severe drought period (Skirycz et al., 2014), which pose severe challenges to plant growth. Environmental legislation in Brazil requires the rehabilitation of disabled iron ore mining sites (Gastauer et al., 2019), which constitute extremely degraded and difficult to restore environments. These mining sites are characterized by extensive vegetation and soil removal, compacted, nutrient-poor soils and steep slopes (Boyer and Wratten, 2010; Garris et al., 2016; Whiting et al., 2004). The successful restoration and rehabilitation of these highly degraded sites thus requires an appropriate selection of plant species (Giannini et al., 2017) and seed sourcing zones to ensure that the introduced plants can effectively colonize and establish viable populations. However, mine rehabilitation programs in the region employ

seed mixtures of exotic plants, due to the scarcity of native seeds and their lower germination and growth rates (Silva et al., 2018). Additionally, some natural Canga environments from the region have been repeatedly disturbed by fires, illegal cattle ranching and the introduction of alien plant species, such as grasses and ferns. This is the case in Serra da Bocaina, which became a National Park in 2017 and has since been protected (Mota et al., 2018). No restoration initiatives have yet been implemented to recover the original ecosystems found in Serra da Bocaina, but provenance strategies are expected to differ substantially from those required to rehabilitate disabled mines.

Being native species, dominant in Canga environments, and once found at the sites being restored, *Mimosa acutistipula var. ferrea* Barneby and *Dioclea apurensis* Kunth (both legumes) are among the most promising plants for use in Canga restoration and mineland rehabilitation programs (Giannini et al., 2017). Considered metallophyte species, both exhibit biological mechanisms to tolerate and thrive in metalliferous soils (Preite et al., 2019; Whiting et al., 2004). Moreover, they are extremely abundant in pristine Cangas ecosystems and interact symbiotically with nitrogen-fixing bacteria, thus contributing to soil enrichment and acting as pioneer species in restoration sites (Nunes et al., 2015; Ramos et al., 2019a; Silva et al., 2018). *Mimosa acutistipula* is drought tolerant and well adapted to the low nutrient content of Canga soils (Silva et al., 2018). On the other hand, *D. apurensis* requires low nutrient inputs and shows high nutrient use efficiency (Ramos et al., 2019b). Moreover, this species is a fast-growing liana with a ground-covering growth form, enabling the revegetation and stabilization of mine pits and waste piles. Both species have high germination rates (Ramos et al., 2019a), can be observed growing on minelands, and seem to be central in plant-pollinator networks (unpublished data). Considering the heterogeneous and hostile

153 environment where both species occur (Mitre et al., 2018), and their life-history similarities,
154 we expected to find similar patterns of neutral and adaptive genetic structure in both species
155 across our study area. We also expected that local populations would not be adapted to the
156 environmental conditions found in exhausted mining sites given they drastically differ from
157 pre-mining conditions, whereas populations from Serra da Bocaina would show adaptations to
158 local environmental conditions. Based on our results, we propose two different site-adjusted
159 provenance strategies for the restoration of a degraded mine site and a disturbed but unmined
160 Canga environment. We discuss the merits of our approach and argue that it can be more
161 broadly applied to define site-adjusted provenancing strategies.
162
163 **Material and Methods**
164
165 *Sampling*
166 We followed a stratified sampling design, seeking to ensure high statistical power in GEA and
167 GPA analyses by maximizing environmental variability within different genetic clusters. We
168 collected samples of 180 individuals of *M. acutistipula* var. *ferrea* and 167 individuals of *D.*
169 *apurensis* between February and May of 2018 (SISBIO collection permit N. 48272-6), across
170 the three major Canga highlands of the Carajás Mineral Province (Fig. 2). For each individual
171 plant, we collected a sample of root-proximal soil (0-5 cm) for chemical characterization and
172 leaflet samples for phenotype and genotype analyses. These Canga ecosystems are composed
173 of several physiognomies, comprising grasslands, scrublands, wetlands and forest formations
174 (Mota et al., 2015), which differ in terms of the plant communities they support as well as in
175 their soil chemistry (Mitre et al., 2018). To ensure sampling across environmental gradients,

individuals were collected in each one of these physiognomies within each highland. We also scattered samples to cover the full extent of each highland (Fig. 2). A minimum distance of 20 meters between samples was used to minimize sampling related individuals. In addition to the soil samples gathered along with plant tissue, we collected 50 extra soil samples from a highly degraded mine site (the mine pit from an exhausted mine) and seven soil samples from a never mined, but moderately disturbed site (scattered across Serra da Bocaina, Fig. 2). These soil samples were used to predict adaptive genotypes associated with the environmental conditions of both restoration sites (see details below).

Environmental data

Soil samples were air-dried and sieved using 2 mm mesh, and once dry they were sent to LABRAS (http://labrasambientaiseagricolas.com.br/) for chemical analyses. These included pH, organic matter, available P, K, and Na, exchangeable Ca, Mg, and Al, exchangeable S and available B, Cu, Fe, Mn and Zn (see details in Supporting Information Methods S1). To reduce the number of chemical parameters describing soil composition, we selected those known to most affect plant physiology in metalliferous ecosystems (organic matter, Fe, Mn, P, pH, S, and B (Bothe, 2011; Mitre et al., 2018; Whiting et al., 2004)), along with a set of orthogonal variables explaining most variation in soil composition across our study area. To identify this set of orthogonal variables we first used the function *imputePCA* from the *missMDA* R package (Josse and Husson, 2016) to impute missing data (20 samples contained missing data for at least one parameter) using the regularized iterative PCA algorithm recommended to avoid overfitting (Josse and Husson, 2016). We then ran separate principal component analyses (PCA) for each species using all the centered and scaled chemical

parameters and selected the three variables showing the strongest correlation with the first, second and third PC axes (each showing eigenvalues > 1 and > 10% of total variance explained, and all three explaining > 50% of total variance; Table S1). The selected soil variables were organic matter, Zn and Na in both species).

We also retrieved climatic data from WorldClim version 1 [1950-2000; (Hijmans et al., 2005)], using the sample coordinates to extract all bioclimatic variables. We followed a similar protocol to obtain the set of orthogonal variables explaining most climatic variance across our study region (Table S1). The first three PCA axes explained 80% of total climatic variance in both species, and the bioclimatic variables most strongly correlated with those axes were isothermality (bio03), minimum temperature of coldest month (bio06) and precipitation of driest quarter (bio17) for *M. acutistipula*; and isothermality (bio03), minimum temperature of coldest month (bio06) and maximum temperature of warmest month (bio05) for *D. apurensis* (see Fig. S1 for maps of these layers). Correlations between these environmental variables were all below $|r| < 0.6$ (Fig. S2).

Phenotypic data

For each leaf sample we determined macro- and micro-nutrient content and the specific leaf area (SLA), using standard methods (see details in Methods S1). We selected those phenotypic variables known to affect plant physiology in metalliferous ecosystems (SLA, N, B, Fe, Mn, P, N/P, (Bothe, 2011; Mitre et al., 2018; Pérez-Harguindeguy et al., 2013). As described above, we also selected the three orthogonal phenotypic variables explaining most of phenotypic variance (the first three PCA axes explained more than 50% of total variance in both species, Table S1) after imputing missing data (11 samples contained missing data in at

222 least one parameter). Selected phenotypic variables were Zn, N and B for *M. acutistipula* and
223 P, Mn and K for *D. apurensis* (correlations between these phenotypic variables were all below
224 $|r| < 0.6$; Fig. S2).
225
226 *Genome size estimation, DNA extraction, genotype-by-sequencing and bioinformatic*
227 *processing*
228 We used flow cytometry to estimate haploid genome size in both species (1C DNA content
229 was 712 Mbp in *M. acutistipula* and 642 Mbp in *D. apurensis,* see details in Methods S1).
230 Total DNA was extracted using Qiagen's DNeasy Plant Mini Kit. DNA concentration was
231 quantified using Qubit High Sensitivity Assay kit (Invitrogen), and DNA integrity assessed
232 through 1.2% agarose gel electrophoresis. Samples with concentrations below 5 ng/µL or
233 showing no clean bands were excluded from all analyses, and selected samples were
234 normalized to a concentration of 5 ng/µL and a total volume of 30 µL. These were then
235 shipped to SNPSaurus (http://snpsaurus.com/) for sequencing and raw data bioinformatic
236 processing (see details in Methods S1). Briefly, genomic DNA was converted into nextRAD
237 genotyping-by-sequencing libraries (SNPsaurus, LLC) as in (Russello et al., 2015),
238 considering the estimated genome size of each species. Genomic DNA was first fragmented
239 with the Nextera DNA Flex reagent (Illumina, Inc), which also ligates short adapter sequences
240 to the ends of the fragments. The Nextera reaction was scaled for fragmenting 14 ng and 20
241 ng of genomic DNA for *M. acutistipula* and *D. apurensis*, respectively. Fragmented DNA was
242 then amplified for 25 cycles at 75 degrees, with one of the primers matching the adapter and
243 extending 8 nucleotides into the genomic DNA with the selective sequence TGCAGGAG.
244 Thus, only fragments starting with a sequence that can be hybridized by the selective

sequence of the primer were efficiently amplified. The nextRAD libraries were then sequenced on a HiSeq 4000 with six and five lanes of 150 bp reads for *M. acutistipula* and *D. apurensis*, respectively (University of Oregon). Reads were trimmed using *BBMap tools* (http://sourceforge.net/projects/bbmap/) to exclude Nextera adapters and a reference contig was created by collecting 10 million reads in total, evenly distributed from the samples, and excluding reads that had counts fewer than 6 or more than 800. The remaining loci were then aligned to each other to identify allele loci and collapse allelic haplotypes to a single representative. All reads were mapped to the reference contig with an alignment identity threshold of 95% using *BBMap tools*. Genotype calling was done using *callvariants* (*BBMap tools*), and the resulting set of genotypes were filtered to remove alleles with population frequency of less than 3%. Loci that were heterozygous in all samples and loci that contained more than 2 alleles in a sample (suggesting collapsed paralogs) were removed. A total of 7,165 RAD-tag sequences were obtained for *M. acutistipula* and 4,325 for *D. apurensis.* Considering the genome size of each species and a linkage block size of 378 Kbp (mean value for the Fabaceae family, (Lowry et al., 2017)), we estimated a maximum proportion of genome coverage (assuming one RAD-tag per block) of 100% (McKinney et al., 2017). From those RAD-tags, 17,403 SNPs were generated for *M. acutistipula* and 9,857 SNPs for *D. apurensis* (minimum sequencing depth of 14 and 9, respectively).

Genetic diversity and neutral genetic structure

The R package *r2vcftools* (https://github.com/nspope/r2vcftools) - a wrapper for VCFtools (Danecek et al., 2011) - was used to perform final quality control on the genotype data. To assess neutral genetic structure and genetic diversity, we used a series of filters to obtain a set

of neutral and independent loci. Filtering criteria included quality (Phred score > 30), read depth (20 – 800), minor allele frequency (MAF > 0.05), linkage disequilibrium (r^2 < 0.8, (Xuereb et al., 2018)), Hardy-Weinberg Equilibrium (HWE, p > 0.0001), and loci and individuals with less than 20% missing data (an example filtering script can be seen in https://github.com/rojaff/r2vcftools_basics). Additionally, we removed loci potentially under selection using genome scans. These accounted for population structure (assessed using the *snmf* function from the *LEA* package, as described below), and controlled for false discovery rates by adjusting *p*-values with the genomic inflation factor (λ) and setting false discovery rates to q=0.05, using the Benjamini-Hochberg algorithm (François et al., 2016) (see details below).

We used two complementary genetic clustering approaches to assess neutral population structure: the *snmf* function from the *LEA* package (Frichot and François, 2015), and Discriminant Analysis of Principal Components - DAPC from the *adegenet* package (Jombart and Ahmed, 2011). The *snmf* model implements a fast yet accurate likelihood algorithm (Frichot et al., 2014), while DAPC is a robust genetic clustering method with no assumption about the underlying population genetic model (Jombart and Ahmed, 2011). Based on previous population genomic studies for other co-occurring plant species (Carvalho et al., 2019; Lanes et al., 2018; Silva et al., 2020), we tested from one to ten ancestral populations (k). In the case of *snmf* we performed ten replicate runs for each value of k, choosing the most likely k based on minimized cross-entropy. For DAPC, we inferred optimal k using k-means clustering and the Bayesian Information Criterion (BIC). Considering the ancestry coefficients assigned by *snmf*, we then estimated expected heterozygosity (H_E), inbreeding coefficients (F), and nucleotide diversity (π) for each genetic cluster. We also

291 estimated pairwise F_{ST} using dartR R package (Gruber et al., 2018), and effective population
292 sizes (N_e) employing the linkage disequilibrium method implemented in NeEstimator 2.1 and
293 a lowest allele frequency value of 0.05 (Do et al., 2014). Finally, we assessed fine-scale
294 spatial genetic structure in each species within each genetic cluster through local polynomial
295 fitting (LOESS) of Yang's genetic relatedness between pairs of individuals (Yang et al.,
296 2010) and pairwise geographic distance, as in (Carvalho et al., 2019).
297
298 *Assessing genotype-environment associations (GEA) and genotype-phenotype associations*
299 *(GPA)*
300 To assess GEA and GPA (Fig. 1) we first filtered loci for quality (Phred score > 30), read
301 depth (20-800), minor allele frequency (MAF > 0.05), linkage disequilibrium ($r^2 < 0.8$), and
302 loci and individuals with less than 20% of missing data. We then combined univariate and
303 multivariate methods, namely Latent Factor Mixed Models (LFMM) and Redundancy
304 Analysis (RDA). While LFMM identifies associations between single loci and single
305 predictors, RDA can detect multilocus signatures of selection as a function of a multivariate
306 set of predictors (Caye et al., 2019; Forester et al., 2018). Both methods assume a linear
307 relationship between allele frequency and environmental variables, have been used
308 extensively (Ahrens et al., 2018), provide a good compromise between detection power and
309 error rates, and are robust to a variety of sampling designs and underlying demographic
310 models (Forester et al., 2018; Rellstab et al., 2015). Since both methods require complete data
311 sets (without missing values), we performed an imputation of missing genotypes (7.6% and
312 7% missing genotypes for *M. acutistipula* and *D. apurensis* respectively) based on the *snmf*
313 population assignments from the previous step, using the *impute* function and the mode

314 method from the *LEA* package (Frichot and François, 2015). This function imputes missing
315 genotypes using ancestry and genotype frequency estimates from the *snmf* run.

316 LFMM analysis were performed using the *lfmm* package (Caye et al., 2019) and ridge
317 estimates, which minimize regularized least-squares with a L_2 penalty (see example script
318 here: https://bcm-uga.github.io/lfmm/articles/lfmm). Instead of using raw predictor variables,
319 we employed the first four axes resulting from a Principal Components Analysis (PCA) on all
320 predictor variables in order to minimize the number of tests. These four axes explained more
321 than 60% of total environmental and phenotypic variance in both species, and were strongly
322 correlated ($|r| > 0.7$) with organic material, B, Fe, Bio06, Bio17, Zn, S and Na (environmental
323 variables), and N/P, P, Fe (phenotypic variables) in both species. We ran LFMM using the
324 previously identified number of genetic clusters (k=3, see results) as latent factors, to account
325 for the underlying neutral genetic structure. We then calculated the genomic inflation factor
326 (λ) and modified it until a calibrated distribution of adjusted p-values was found, and set false
327 discovery rates at a rate of q=0.05 using the Benjamini–Hochberg algorithm (François et al.,
328 2016).

329 We performed RDA using the *rda* function from the *vegan* package (Oksanen et al.,
330 2019) as implemented in Forester et al. (2018), modeling genotypes as a function of predictor
331 variables, and producing as many constrained axes as predictors (see example script here:
332 https://popgen.nescent.org/2018-03-27_RDA_GEA.html). Multicollinearity between
333 predictors was assessed using the variance inflation factor (VIF) and since all predictor
334 variables showed VIF < 3 none were excluded. Raw predictor variables were scaled and
335 centered prior to analyses and the population assignments from *snmf* (population ID) were
336 used to control for population structure by running a partial RDA. Significance of RDA

337 constrained axes was assessed using the *anova.cca* function and significant axes were then
338 used to identify candidate loci in both species. Candidate loci were identified using a
339 Mahalanobis distance-based approach (Capblancq et al., 2018), which made RDA result
340 comparable with those obtained with LFMM, since it allowed adjusting *p*-values using the
341 genomic inflation factor (λ) and setting false discovery rates to q=0.05, as described above
342 (calculated and modified genomic inflation factors and *p*-value distributions for LFMM and
343 RDA tests are provided in Figs. S3-S8). To assess the impact of population genetic structure
344 on our number of detections we ran additional cluster-level GEA analyses (LFMM and RDA),
345 using only individuals belonging to the same genetic cluster (setting k=1 in LFMM and
346 omitting population ID in RDA). Finally, to visualize patterns of GEA and GPA, we ran
347 additional RDA models excluding neutral loci, using the combined candidate adaptive loci
348 detected using the general RDA and LFMM analyses.

349 In order to search for the proteins coded by the genes contained in the flanking regions
350 of our candidate SNPs (found in GEA and GPA analyses), contig sequences containing
351 candidate loci were first submitted to the EMBOSS Transeq
352 (http://www.ebi.ac.uk/Tools/st/emboss_transeq/) to obtain corresponding protein sequences.
353 We used all six frames with standard code (codon table), regions (start-end), trimming (yes),
354 and reverse (no). We then ran a functional analysis using InterPro
355 (https://www.ebi.ac.uk/interpro/; interproscan.sh -dp –appl PfamA, TIGRFAM, PRINTS,
356 PrositePatterns, Gene3d –goterms –pathways -f tsv -o MySequences.tsv -i MySequences.faa),
357 searching for gene ontology terms and pathways along a variety of annotation databases (i.e.,
358 Interpro, Pfam, Tigrfam, Prints, PrositePattern and Gene3d).
359

Mapping adaptive genetic variation

To map adaptive genetic variation, we used the *adegenet* package (Jombart and Ahmed, 2011) to run a Spatial Principal Component Analysis (sPCA) on the combined candidate adaptive loci detected in GEA and GPA analyses using general LFMM and RDA (results for intersected loci are presented in Fig. S15). sPCA is a spatially explicit multivariate method that yields scores summarizing genetic variability and spatial structure among individuals (Jombart et al., 2008). Spatial structure is estimated using a Moran's Index that relies on the comparison of allelic frequencies observed in one individual to the values observed in neighboring individuals. These neighboring individuals can be defined by distinct connection networks, which in our case was set to a distance-based neighborhood, as indicated for aggregated distributions (Jombart et al., 2008). The Moran's Index generates two types of spatial structuring: global structure, which reflects positive spatial autocorrelation, and local structure, that reflects negative spatial autocorrelation (Jombart et al., 2008). To decide if global and/or local structures should be interpreted and thus retained in sPCA analyses, we used the global and local tests proposed by Jombart & Ahmed (2011). The first three retained axes were then interpolated on 10 meter resolution grids covering our study area, and the resulting rasters used to create an RGB composite, using the Merge function in QGIS 3.4 (see example scripts here: https://github.com/rojaff/LanGen_pipeline). The resulting color patterns represent the similarity in adaptive genetic composition.

To predict the adaptive genotypes associated with environmental data collected from restoration sites (the highly degraded exhausted mine and the moderately disturbed Serra da Bocaina, Fig. 2), we employed the GEA-RDA models fitted on the combined candidate adaptive loci detected by global LFMM and RDA (see previous section), and ran the

383 *predict.cca* function from the *vegan* package. Environmental samples (soil and climate) from
384 these sites were thus used to predict RDA scores, based on the fitted GEA-RDA models. We
385 then performed a *k*-means clustering analysis (using Euclidean distances) on observed and
386 predicted RDA scores for individuals from each species, using all significant constrained axes
387 and allowing the number of clusters to vary between two and five (three Canga highlands and
388 two restoration sites). We used the *NbClust* package (Charrad et al., 2014) to obtain the
389 optimal number of clusters chosen by 30 different algorithms. Observed and predicted RDA
390 scores groupping together, suggest that our sampled individuals possess adaptations
391 associated with the environmental conditions of restoration sites. Observed and predicted
392 RDA scores placed in different clusters, on the other hand, indicate that none of our sampled
393 individuals seems adapted to the environmental conditions of restoration sites.
394
395 **Results**
396
397 *Genetic diversity and neutral genetic structure*
398 After filtering for quality, read depth, minor allele frequencies, missing data, linkage
399 disequilibrium, Hardy-Weinberg Equilibrium, and outlier loci, we retained 7,376 and 3,496
400 neutral and independent SNPs and 177 and 163 individuals for *M. acutistipula* and *D.*
401 *apurensis*, respectively, which were then used to assess genetic diversity and population
402 structure. Both genetic clustering approaches (*snmf* and DAPC) indicated the presence of
403 three clusters in the two study species (Fig. S9). Admixture levels were low, all individuals
404 were correctly assigned to their source Canga highland (Fig. 2), and there was genetic
405 differentiation between genetic clusters (pairwise F_{ST} values were significant and ranged

406 between 0.11 and 0.13 in *M. acutistipula* and between 0.16 and 0.27 in *D. apurensis*).
407 Expected heterozygosity and nucleotide diversity were similar in both species, but inbreeding
408 coefficients were lower and effective population sizes larger in *M. acutistipula* (Table 1). Both
409 species showed significant inbreeding coefficients in all genetic clusters and exhibited the
410 largest effective population sizes in Serra Sul (Table 1). We detected spatial autocorrelation in
411 genetic relatedness within genetic clusters in each species (Fig. S10-S11). In both, the strength
412 of spatial autocorrelation was highest in Serra Sul, where genetic neighborhood size was
413 larger (~5km, Fig. S10-S11).
414
415 *Genotype-environment and genotype-phenotype associations*
416 After filtering for quality, read depth, minor allele frequencies, missing data, and linkage
417 disequilibrium we retained 9,480 and 4,720 SNPs and 177 and 163 individuals for *M.*
418 *acutistipula* and *D. apurensis*, respectively. Using LFMM we identified a total of 198 and 154
419 contigs (RAD-tags) containing GEA, and 94 and 185 contigs containing GPA in *M.*
420 *acutistipula* and *D. apurensis*, respectively (Tables S2 and S3). Only the first two constrained
421 axes from RDA analyses were significant (ANOVA's $p < 0.05$) in GEA and GPA analyses for
422 both species. RDA revealed a total of 403 and 225 contigs containing significant GEA and
423 281 and 119 contigs containing significant GPA in *M. acutistipula* and *D. apurensis*
424 respectively (Fig. 3, Fig. S12 and Tables S2 and S3). In *M. acutistipula* 344 contigs were most
425 correlated to climatic variables and 69 to soil variables, while in *D. apurensis* 203 contings
426 were most correlated to climatic and 23 to soil variables. Combining both methods (LFMM
427 and RDA), we found a total of 588 contigs showing GEA in *M. acutistipula* and 360 in *D.*
428 *apurensis*, and 368 contigs showing GPA in *M. acutistipula* and 288 in *D. apurensis*. Only

108 contigs contained both GEA and GPA in *M. acutistipula* and 65 in *D. apurensis*. Finally, cluster-level GEA analyses revealed many cluster-exclusive detections in both species (Fig. S13).

Subsequent RDA models using the combined candidate adaptive loci detected using general LFMM and RDA analyses revealed population-level patterns of GEA and GPA (Fig. 4). In *M. acutistipula* GEA and GPA models explained 17% and 5% of total variance respectively, while in *D. apurensis* GEA and GPA models explained 31% and 9% of total variance. In both species, axes loadings were higher for climatic variables (0.01-0.89 for *M. acutistipula* and 0.01-0.83 for *D. apurensis*) than for soil variables (0.005-0.47 for *M. acutistipula* and 0.003-0.59 for *D. apurensis*). In *M. acutistipula*, the first and second axes split individuals into three large GEA groups corresponding to their sampling location. While individuals from Serra Norte showed associations with higher isothermality (bio03) and higher winter temperatures (bio06), individuals from Serra Sul showed associations with warmer winter temperatures and wetter dry season precipitation (bio17). Individuals from Serra da Bocaina exhibited associations with higher pH (less acidic soils) and drier dry season precipitation (Fig. 4a). Interestingly, *Dioclea apurensis* showed similar GEA patterns based on isothermality (bio03), winter temperatures (bio06), and pH, despite using a slightly different set of predictors (Fig. 4b). On the other hand, the first and second constrained axes divided individuals into two large GPA groups in *M. acutistipula* (Fig. 4c), the first one encompassing individuals from Serra Norte (which showed associations with higher SLA and Mn, and lower P), and the second individuals from Serra Sul and Serra da Bocaina (showing associations with lower SLA and Mn). In *D. apurensis*, the first and second axes split individuals into three GPA groups, with individuals from Serra Norte showing associations with a higher leaf

content of Fe and Mn and a lower content of P, while those from Serra Sul showed associations with higher N and those from Serra da Bocaina with lower SLA and N/P (Fig. 4d). Leaf-level nutrients were weakly correlated with soil-level nutrients (Pearson's correlation coefficients ranged between -0.07 and 0.39 for *M. acutistipula* and between -0.04 to 0.24 for *D. apurensis*).

A subset of the contigs containing candidate SNPs showed InterPro annotations (105 contigs in *M. acutistipula* and 59 in *D. apurensis*). Candidate adaptive genes were associated to different functions, including intracellular transport, catalytic activity, synthesis of hormones, metabolic and oxidation-reduction processes, and plant defense response (a full list of candidate genes with InterPro annotations is presented in Table S4). Only 17 putative adaptive genes containing InterPro annotations were shared between both species (Table S5).

Mapping adaptive genetic variation

The combination (union) of candidate adaptive loci detected through GEA and GPA resulted in 914 loci for *M. acutistipula* and 614 loci for *D. apurensis.* Since none of the sPCA local structure tests were significant, we retained the first three positive global axes, which explained most variance in both species (51% and 81% of total variance for *M. acutistipula* and *D. apurensis* respectively, Fig. S14). These revealed a similar adaptive genetic structure in both species (Fig. 5), with two adaptive units in Serra Norte and one in Serra da Bocaina. *Mimosa acutistipula* nevertheless exhibited a clinal adaptive pattern in Serra Sul, whereas *D. apurensis* did not. Similar spatial patterns were found when using the intersected loci (i.e. those shared by GEA and GPA; Fig. S15). Finally, predicted genotypes associated with climatic and soil characteristics from a highly degraded mining site did not cluster together

475 with any of our study populations in either species (Fig. 6a and 6b). In contrast, most
476 predicted genotypes for the environmental conditions from the moderately disturbed Serra da
477 Bocaina clustered together with individuals collected in the same location, revealing local
478 provenances are putatively adapted to local environmental conditions (Fig. 6c and 6d).
479
480 **Discussion**
481
482 The delineation of seed sourcing areas requires accounting for evolutionary history, genetic
483 diversity, and how likely individuals will adapt to the environmental conditions of the targeted
484 restoration sites (Breed et al., 2019). Here we employ a comprehensive landscape genomic
485 approach to characterize neutral and adaptive genetic variation and provide insights to assist
486 the restoration of a highly degraded mining site and a moderately disturbed Canga highland
487 from the Carajás Mineral Province. We discuss how our results can help define site-adjusted
488 provenancing strategies and argue that our methods can be more broadly applied to assist
489 other restoration and rehabilitation initiatives.
490 Several studies have stressed the importance of avoiding inbreeding, increasing
491 genetic diversity to maintain evolutionary potential, and minimizing outbreeding depression
492 in restored populations (Broadhurst et al., 2008; Hufbauer et al., 2015; Mijangos et al., 2015;
493 Weeks et al., 2011). The assessment of neutral genetic structure provides information on how
494 to minimize outbreeding depression by avoiding mixing individuals from different
495 evolutionary lineages (Mijangos et al., 2015). Estimates of the genetic neighborhood size, on
496 the other hand, provide clues on how to sample unrelated individuals within seed sourcing
497 areas to increase genetic diversity and reduce the risk of inbreeding depression in restored

498 populations (Breed et al., 2019; Krauss and Koch, 2004). Our initial assessment of neutral
499 genetic structure identified three demographically independent units (or Management Units
500 *sensu* (Funk et al., 2012)), which could be considered distinct provenances to minimize the
501 risk of outbreeding depression (Frankham et al., 2017). Within these zones, our estimates of
502 genetic neighborhood size provide information on within-cluster seed sourcing strategies to
503 maximize genetic diversity. In Serra Sul, for example, seed sources located 5 Km apart are not
504 expected to be related (Fig. S10-S11), and would comprise a better representation of standing
505 genetic variation than individuals collected across smaller spatial scales. Effective population
506 size estimates (Table 1) nevertheless indicate that none of our observed genetic clusters is
507 likely to experience inbreeding depression in the near future based on the 50/500 rule
508 (Jamieson and Allendorf, 2012). Inbreeding levels observed in both species were nonetheless
509 significantly different from zero, suggesting some level of selfing or mating between related
510 individuals is taking place.

511 Patterns of local adaptation will ultimately determine the ability of plants to
512 effectively colonize and quickly recover disturbed sites (Mijangos et al., 2015). By using the
513 combined candidate loci detected in GEA and GPA, using both univariate and multivariate
514 methods, we improved the detection of single-locus and multi-locus adaptive signals.
515 (Mahony et al., 2019; Talbot et al., 2016; Vangestel et al., 2018). Interestingly, more
516 intersections between GEA and GPA were found when using RDA than when using LFMM
517 (Fig. 3), indicating that most adaptations to local environmental conditions expressing
518 differential phenotypes are polygenic (Forester et al., 2018). Indeed most fitness-related traits
519 in plants have a polygenic basis (Falke et al., 2013), including tolerance to soil with
520 phytotoxic levels of heavy metals (Arnold et al., 2016). We nevertheless note that other genes

521 occurring in the flanking regions of our candidate SNPs could be responsible for the detected
522 adaptive signals, and that many sequences did not match translated proteins, or found matches
523 with uncharacterized proteins. Still, the most frequent amongst our identified candidate
524 proteins are involved in plant defense and stress responses (including reverse transcriptase,
525 ribonuclease H-like domain, P-loop NTPase fold, leucine-rich repeat, and thaumatin) or basic
526 metabolic processes (pentatricopeptide repeat, protein kinase domain, Nitrite/Sulfite reductase
527 ferredoxin-like domain, major intrinsic protein, and kinesin motor domain), suggestive of
528 adaptations to harsh environments (Tables S4 and S5). Some of these putative adaptive genes
529 were shared between the two species as well as with other co-occurring species (Table S5),
530 indicating convergent evolution to similar environmental pressures (Arnold et al., 2016;
531 Yeaman et al., 2016). These shared genes thus constitute primary targets for functional studies
532 investigating the molecular basis of adaptation to Canga environments and minelands.

533 Cluster-level GEA analyses revealed many cluster-exclusive detections in both species
534 (Fig. S13), suggesting that microgeographic adaptation may play a role in driving genetic
535 patterns within highlands (Richardson et al., 2014). To visualize and better understand the
536 mechanisms behind the observed GEA and GPA we ran additional RDAs using the combined
537 candidate adaptive loci detected in our general LFMM and RDA analyses. As expected, we
538 found similar patterns of GEA in both species (Figs. 4a and 4b). Interestingly, the strongest
539 GEA were found with climatic variables in both species, in spite of the coarse resolution of
540 WorldClim data and the narrow climatic variation found across our study area (Fig. S1). Our
541 results thus suggest that local climate constitutes an important environmental filter driving
542 local adaptation, as found in other species from Canga environments (Lanes et al., 2018) and
543 temperate climates (Pais et al., 2017; Pluess et al., 2016). In *M. acutistipula,* Serra Norte

544 populations showed associations with higher SLA, suggesting climatic or soil conditions in
545 Serra Norte are more favorable to plant growth (He et al., 2018). Individuals from Serra da
546 Bocaina and Serra Sul showed associations with lower SLA and lower concentration of
547 several micro- and macro-nutrients, suggesting that increasing leaf thickness in these
548 individuals avoids dissection or better preserves scarce nutrients (Costa-Saura et al., 2016). In
549 contrast, SLA-associations in *D. apurensis* did not separate Canga highlands, showing that the
550 influence of climatic variation on SLA is different across species (Gong and Gao, 2019; Liu et
551 al., 2017). In *D. apurensis*, different genotype associations with leaf micro and macronutrients
552 separated highlands (Fig. 4d), suggesting different physiological requirements or nutrient
553 availability at each site. Low correlations between leaf and soil Fe and Mn concentrations,
554 suggest our study species are controlling nutrient absorption, which makes them suitable for
555 the restoration of areas with a high concentration of these metals. Controlled common-garden
556 or reciprocal transplant experiments are nevertheless needed to assess growth and overall
557 performance of different genotypes (sources) in different soils and climates (Aitken and
558 Bemmels, 2016; Rellstab et al., 2015).

559 Our local adaptation maps reveal areas containing similar local adaptations (colors) in
560 each species (Fig. 5, Fig. S15), which could be used to delineate seed sourcing strategies. In
561 contrast to the commonly employed Generalized Dissimilarity Models (GDM) (Gugger et al.,
562 2018; Rossetto et al., 2019; Shryock et al., 2015; Supple et al., 2018), our mapping approach
563 based on sPCA allows incorporating GPA and predicting adaptive genetic variation from site-
564 level data, which is particularly useful for areas lacking high-resolution environmental layers.
565 Moreover, sPCA explicitly account for spatial autocorrelation in genetic composition, which
566 is likely to play an important role explaining patterns of local adaptation (Lesica and

567 Allendorf, 1999; Richardson et al., 2014) (see Fig. S16 for alternative adaptation maps
568 generated using GDM). Our adaptation maps showed similar adaptations across Serra da
569 Bocaina (i.e. a single adaptive unit), and most predicted genotypes associated with local
570 environmental samples (climate and soil) clustered together with individuals sampled in Serra
571 da Bocaina. This result indicates that local provenances are probably best adapted to local
572 environmental conditions at this site under contemporary climates (Fig. 6c and 6d), and
573 supports the recommendations made by Lesica & Allendorf (1999) for the restoration of
574 moderately disturbed sites. Since genetic neighborhood size in Serra da Bocaina was roughly
575 3 Km for both species, our results suggest that local seeds collected in areas separated by at
576 least 3 Km would maximize genetic diversity at this location.

577 In contrast, predicted genotypes for the environmental data collected at the degraded
578 mine site did not cluster with any of our study populations in either species (Fig. 6a and 6b).
579 This indicates that none of the genotypes we sampled from natural habitats overlap with the
580 multivariate environmental conditions present at the mine site. In this case, mixing genotypes
581 containing different local adaptations could be regarded as the best option to maximize
582 evolutionary potential and facilitate adaptation to novel environments (Lesica and Allendorf,
583 1999). Seeds could be sourced from all the identified adaptive units (colors in sPCA maps);
584 and within these units they could be sampled in areas separated by the genetic neighborhood
585 size to further enhance genetic diversity. Although mixing individuals from different
586 management units could result in outbreeding depression (Hufford and Mazer, 2003; Weeks et
587 al., 2011), the risk is likely marginal for these study species, which are widely distributed
588 across the continent (Dutra and Morim, 2015; Queiroz, 2015). Moreover, environmental
589 conditions and plant communities show remarkable similarities across the Carajás Mineral

590 Province when compared to other *campo rupestre* formations (Zappi et al., 2019). Such a
591 regional admixture provenancing approach (Bucharova et al., 2019) would represent a
592 significant improvement over introducing exotic species, which are currently being used in
593 mine reclamation programs due to the availability of seeds in large quantities and their ability
594 to quickly colonize mine environments to prevent soil erosion (D'Antonio and Meyerson,
595 2002; Gastauer et al., 2019; Silva et al., 2018). Our results could guide the establishment of
596 seed production areas for both native species, aiming to overcome shortfalls in seed
597 availability while capturing standing neutral and adaptive genetic variation (Nevill et al.,
598 2016).

599 Our work illustrates how neutral and adaptive genetic variation can be used to provide
600 evidence-based recommendations for provenance schemes aimed to effectively restore sites
601 ranging between moderately disturbed and highly degraded. In our two study species, local
602 provenances were found optimal to restore a moderately disturbed site, whereas a mixture of
603 genotypes was suggested as the most promising strategy to recover a highly degraded mining
604 site, to which local provenances were not adapted. Our proposed methodological approach
605 (Fig. 1) can be more broadly applied to define site-adjusted provenance strategies in other
606 locations and for other disturbance regimes. We recognize that the high costs associated with
607 genomic analyses and the complexity of bioinformatic and statistical analyses represent
608 important barriers for practitioners (Breed et al., 2019; Shafer et al., 2015). Still, as genomic
609 data becomes available for more species exhibiting different life-history characteristics,
610 restoration genomic initiatives using similar methods coupled with visually appealing and
611 user-friendly interfaces (Rossetto et al., 2019), are likely to substantially improve restoration
612 outcomes.

Acknowledgments

Funding was provided by Instituto Tecnológico Vale, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grants 301616/2017-5 (RJ), 153535/2018-0 (MG) and 316067/2018-0 (LCRM), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) grant 88887.156652/2017-00 (CSC). We thank Cesar Neto and Eder Lanes for assistance in the field, Nelson Carvalho and Santelmo Vasconcelos for assistance in genome size estimation, Eder Lanes and Manoel Lopes for help in the laboratory, and three anonymous referees for improving earlier versions of this manuscript.

References

Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, Dudaniec RY. 2018. The search for loci under selection: trends, biases and progress. *Mol Ecol* **27**:1342–1356. doi:10.1111/mec.14549

Aitken SN, Bemmels JB. 2016. Time to get moving: Assisted gene flow of forest trees. *Evol Appl* **9**:271–290. doi:10.1111/eva.12293

Allendorf FW, Luikart GH, Aitken SN. 2013. Conservation and the genetics of populations. Oxford: Wiley-Blackwell.

Arnold BJ, Lahner B, DaCosta JM, Weisman CM, Hollister JD, Salt DE, Bomblies K, Yant L. 2016. Borrowed alleles and convergence in serpentine adaptation. *Proc Natl Acad Sci* **113**:8320–8325. doi:10.1073/pnas.1600405113

Balkenhol N, Dudaniec RY, Krutovsky K V, Johnson JS, Cairns DM, Segelbacher G, Selkoe KA, von der Heyden S, Wang IJ, Selmoni O, Joost S. 2017. Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations. Cham: Springer International Publishing. pp. 261–322. doi:10.1007/13836_2017_2

Bothe H. 2011. Plants in Heavy Metal SoilsDetoxification of Heavy Metals. pp. 35–57. doi:10.1007/978-3-642-21408-0_2

642 Boyer S, Wratten SD. 2010. The potential of earthworms to restore ecosystem services after
643 opencast mining - A review. *Basic Appl Ecol* **11**:196–203.
644 doi:10.1016/j.baae.2009.12.005

645 Breed MF, Harrison PA, Blyth C, Byrne M, Gaget V, Gellie NJC, Groom SVC, Hodgson R,
646 Mills JG, Prowse TAA, Steane DA, Mohr JJ. 2019. The potential of genomics for
647 restoring ecosystems and biodiversity. *Nat Rev Genet*. doi:10.1038/s41576-019-0152-0

648 Breed MF, Stead MG, Ottewell KM, Gardner MG, Lowe AJ. 2013. Which provenance and
649 where? Seed sourcing strategies for revegetation in a changing environment. *Conserv*
650 *Genet* **14**:1–10. doi:10.1007/s10592-012-0425-z

651 Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Vesk PA, Yates C. 2008.
652 Seed supply for broadscale restoration: Maximizing evolutionary potential. *Evol Appl*
653 **1**:587–597. doi:10.1111/j.1752-4571.2008.00045.x

654 Bucharova A, Bossdorf O, Hölzel N, Kollmann J, Prasse R, Durka W. 2019. Mix and match:
655 regional admixture provenancing strikes a balance among different seed-sourcing
656 strategies for ecological restoration. *Conserv Genet* **20**:7–17. doi:10.1007/s10592-018-
657 1067-6

658 Capblancq T, Luu K, Blum MGB, Bazin E. 2018. Evaluation of redundancy analysis to
659 identify signatures of local adaptation. *Mol Ecol Resour* **18**:1223–1233.
660 doi:10.1111/1755-0998.12906

661 Carvalho CS, Lanes ÉCM, Silva AR, Caldeira CF, Carvalho-Filho N, Gastauer M, Imperatriz-
662 Fonseca VL, Nascimento Júnior W, Oliveira G, Siqueira JO, Viana PL, Jaffé R. 2019.
663 Habitat Loss Does Not Always Entail Negative Genetic Consequences. *Front Genet* **10**.
664 doi:10.3389/fgene.2019.01101

665 Caye K, Jumentier B, Lepeule J, François O. 2019. LFMM 2: fast and accurate inference of
666 gene-environment associations in genome-wide studies. *Mol Biol Evol* **36**:852–860.

667 Charrad M, Ghazzali N, Boiteux V, Niknafs A. 2014. NbClust: An R Package for Determining
668 the Relevant Number of Clusters in a Data Set. *J Stat Softw* **61**:1–36.

669 Coates DJ, Byrne M, Moritz C. 2018. Genetic Diversity and Conservation Units: Dealing
670 With the Species-Population Continuum in the Age of Genomics. *Front Ecol Evol* **6**:1–
671 13. doi:10.3389/fevo.2018.00165

672 Costa-Saura JM, Martínez-Vilalta J, Trabucco A, Spano D, Mereu S. 2016. Specific leaf area
673 and hydraulic traits explain niche segregation along an aridity gradient in Mediterranean
674 woody species. *Perspect Plant Ecol Evol Syst* **21**:23–30.
675 doi:10.1016/j.ppees.2016.05.001

D'Antonio C, Meyerson LA. 2002. Exotic Plant Species as Problems and Solutions in Ecological Restoration: A Synthesis. *Restor Ecol* **10**:703–713. doi:10.1046/j.1526-100X.2002.01051.x

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group 1000 Genomes Project Analysis. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. doi:10.1093/bioinformatics/btr330

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Resour* **14**:209–214. doi:10.1111/1755-0998.12157

Durka W, Michalski SG, Berendzen KW, Bossdorf O, Bucharova A, Hermann J-M, Hölzel N, Kollmann J. 2017. Genetic differentiation within multiple common grassland plants supports seed transfer zones for ecological restoration. *J Appl Ecol* **54**:116–126. doi:10.1111/1365-2664.12636

Dutra VF, Morim MP. 2015. Mimosa in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro. *http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB116253*.

Falke KC, Glander S, He F, Hu J, de Meaux J, Schmitz G. 2013. The spectrum of mutations controlling complex traits and the genetics of fitness in plants. *Curr Opin Genet Dev* **23**:665–671. doi:10.1016/j.gde.2013.10.006

Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol Ecol* **27**:2215–2233. doi:10.1111/mec.14584

François O, Martins H, Caye K, Schoville SD. 2016. Controlling false discoveries in genome scans for selection. *Mol Ecol* **25**:454–469.

Frankham R, Ballou JD, Ralls K, Eldridge M, Dudash MR, Fenster CB, Lacy RC, Sunnucks P. 2017. Genetic management of fragmented animal and plant populations. Oxford University Press.

Frichot E, François O. 2015. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol* **6**:925–929. doi:10.1111/2041-210X.12382

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. 2014. Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics* **196**:973–983. doi:10.1534/genetics.113.160572

708 Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. 2012. Harnessing genomics for
709 delineating conservation units. *Trends Ecol Evol* **27**:489–496.
710 doi:http://dx.doi.org/10.1016/j.tree.2012.05.012

711 Garris HW, Baldwin SA, Van Hamme JD, Gardner WC, Fraser LH. 2016. Genomics to assist
712 mine reclamation: A review. *Restor Ecol* **24**:165–173. doi:10.1111/rec.12322

713 Gastauer M, Souza Filho PWM, Ramos SJ, Caldeira CF, Silva JR, Siqueira JO, Furtini Neto
714 AE. 2019. Mine land rehabilitation in Brazil: Goals and techniques in the context of legal
715 requirements. *Ambio* **48**:74–88. doi:10.1007/s13280-018-1053-8

716 Giannini TC, Giulietti AM, Harley RM, Viana PL, Jaffe R, Alves R, Pinto CE, Mota NFO,
717 Caldeira Jr CF, Imperatriz-Fonseca VL, Furtini AE, Siqueira JO. 2017. Selecting plant
718 species for practical restoration of degraded lands using a multiple-trait approach.
719 *Austral Ecol* **42**:510–521. doi:10.1111/aec.12470

720 Gong H, Gao J. 2019. Soil and climatic drivers of plant SLA (specific leaf area). *Glob Ecol*
721 *Conserv* **20**:e00696. doi:10.1016/j.gecco.2019.e00696

722 Gruber B, Unmack PJ, Berry OF, Georges A. 2018. dartr: An r package to facilitate analysis
723 of SNP data generated from reduced representation genome sequencing. *Mol Ecol*
724 *Resour* **18**:691–699.

725 Gugger PF, Liang CT, Sork VL, Hodgskiss P, Wright JW. 2018. Applying landscape genomic
726 tools to forest management and restoration of Hawaiian koa (Acacia koa) in a changing
727 environment. *Evol Appl* **11**:231–242. doi:10.1111/eva.12534

728 He D, Chen Y, Zhao K, Cornelissen JHC, Chu C. 2018. Intra- and interspecific trait variations
729 reveal functional relationships between specific leaf area and soil niche within a
730 subtropical forest. *Ann Bot* **121**:1173–1182. doi:10.1093/aob/mcx222

731 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution
732 interpolated climate surfaces for global land areas. *Int J Climatol* **25**:1965–1978.
733 doi:10.1002/joc.1276

734 Hufbauer RA, Szucs M, Kasyon E, Youngberg C, Koontz MJ, Richards C, Tuff T, Melbourne
735 BA. 2015. Three types of rescue can avert extinction in a changing environment. *Proc*
736 *Natl Acad Sci* **112**:10557–10562. doi:10.1073/pnas.1504732112

737 Hufford KM, Mazer SJ. 2003. Plant ecotypes: Genetic differentiation in the age of ecological
738 restoration. *Trends Ecol Evol* **18**:147–155. doi:10.1016/S0169-5347(03)00002-8

739 Jamieson IG, Allendorf FW. 2012. How does the 50 / 500 rule apply to MVPs? *Trends Ecol*
740 *Evol* **27**:578–584. doi:10.1016/j.tree.2012.07.001

741 Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
742 data. *Bioinformatics* **27**:3070–3071.

743 Jombart T, Devillard S, Dufour A-B, Pontier D. 2008. Revealing cryptic spatial patterns in
744 genetic variability by a new multivariate method. *Heredity (Edinb)* **101**:92–103.
745 doi:10.1038/hdy.2008.34

746 Josse J, Husson F. 2016. missMDA: A Package for Handling Missing Values in Multivariate
747 Data Analysis. *J Stat Softw* **70**:1–31. doi:10.18637/jss.v070.i01

748 Krauss SL, Koch JM. 2004. Rapid genetic delineation of provenance for plant community
749 restoration. *J Appl Ecol* **41**:1162–1173. doi:10.1111/j.0021-8901.2004.00961.x

750 Krauss SL, Sinclair EA, Bussell JD, Hobbs RJ. 2013. An ecological genetic delineation of
751 local seed-source provenance for ecological restoration. *Ecol Evol* **3**:2138–2149.
752 doi:10.1002/ece3.595

753 Lanes ÉC, Pope NS, Alves R, Carvalho Filho NM, Giannini TC, Giulietti AM, Imperatriz-
754 Fonseca VL, Monteiro W, Oliveira G, Silva AR, Siqueira JO, Souza-Filho PW,
755 Vasconcelos S, Jaffé R. 2018. Landscape Genomic Conservation Assessment of a
756 Narrow-Endemic and a Widespread Morning Glory From Amazonian Savannas. *Front*
757 *Plant Sci* **9**:532. doi:10.3389/fpls.2018.00532

758 Lesica P, Allendorf FW. 1999. Ecological Genetics and the Restoration of Plant Communities:
759 Mix or Match? *Restor Ecol* **7**:42–50. doi:10.1046/j.1526-100X.1999.07105.x

760 Liu M, Wang Z, Li S, Lü X, Wang X, Han X. 2017. Changes in specific leaf area of dominant
761 plants in temperate grasslands along a 2500-km transect in northern China. *Sci Rep*
762 **7**:10780. doi:10.1038/s41598-017-11133-z

763 Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2017.
764 Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing
765 for genome scans of adaptation. *Mol Ecol Resour* **17**:142–152. doi:10.1111/1755-
766 0998.12635

767 Lu M, Krutovsky K V, Loopstra CA. 2019. Predicting Adaptive Genetic Variation of Loblolly
768 Pine (Pinus taeda L.) Populations Under Projected Future Climates Based on
769 Multivariate Models. *J Hered* **110**:857–865. doi:10.1093/jhered/esz065

770 Mahony CR, MacLachlan IR, Lind BM, Yoder JB, Wang T, Aitken SN. 2019. Evaluating
771 genomic data for management of local adaptation in a changing climate: A lodgepole
772 pine case study. *Evol Appl* eva.12871. doi:10.1111/eva.12871

773 Martins K, Gugger PF, Llanderal-Mendoza J, González-Rodríguez A, Fitz-Gibbon ST, Zhao
774 JL, Rodríguez-Correa H, Oyama K, Sork VL. 2018. Landscape genomics provides

775 evidence of climate-associated genetic variation in Mexican populations of Quercus
776 rugosa. *Evol Appl* **11**:1842–1858. doi:10.1111/eva.12684

777 McKinney GJ, Larson WA, Seeb LW, Seeb JE. 2017. RADseq provides unprecedented
778 insights into molecular ecology and evolutionary genetics: comment on Breaking RAD
779 by Lowry et al . (2016). *Mol Ecol Resour* **17**:356–361. doi:10.1111/1755-0998.12649

780 Mijangos JL, Pacioni C, Spencer PBS, Craig MD. 2015. Contribution of genetics to
781 ecological restoration. *Mol Ecol* **24**:22–37. doi:10.1111/mec.12995

782 Mitre SK, Mardegan SF, Caldeira CF, Ramos SJ, Furtini Neto AE, Siqueira JO, Gastauer M.
783 2018. Nutrient and water dynamics of Amazonian canga vegetation differ among
784 physiognomies and from those of other neotropical ecosystems. *Plant Ecol* **219**:1341–
785 1353. doi:10.1007/s11258-018-0883-6

786 Mota NF de O, Silva LVC, Martins FD, Viana PL. 2015. Vegetação sobre sistemas
787 ferruginosos da Serra dos Carajás. *Geossistemas Ferruginosos no Bras Inst Pr{\'\i}stino,*
788 *Belo Horiz* 289–315.

789 Mota NF de O, Watanabe MTC, Zappi DC, Hiura AL, Pallos J, Viveros RS, Giulietti AM,
790 Viana PL. 2018. Cangas da Amazônia: a vegetação única de Carajás evidenciada pela
791 lista de fanerógamas. *Rodriguésia* **69**:1435–1488. doi:10.1590/2175-7860201869336

792 Nevill PG, Tomlinson S, Elliott CP, Espeland EK, Dixon KW, Merritt DJ. 2016. Seed
793 production areas for the global restoration challenge. *Ecol Evol* **6**:7490–7497.
794 doi:10.1002/ece3.2455

795 Nunes JA, Schaefer CEGR, Júnior WGF, Neri A V., Correa GR, Enright NJ. 2015. Soil-
796 vegetation relationships on a banded ironstone 'island', Carajás Plateau, Brazilian
797 Eastern Amazonia. *An Acad Bras Cienc* **87**:2097–2110. doi:10.1590/0001-
798 376520152014-0106

799 Oksanen J, Blanchet F, Friendly, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R,
800 Simpson G, Solymos P, Stevens M, Szoecs E, Wagner H. 2019. vegan: Community
801 Ecology Package.

802 Pais AL, Whetten RW, Xiang Q-YJ. 2017. Ecological genomics of local adaptation in Cornus
803 florida L. by genotyping by sequencing. *Ecol Evol* **7**:441–465. doi:10.1002/ece3.2623

804 Pérez-Harguindeguy N, Diaz S, Gamier E, Lavorel S, Poorter H, Jaureguiberry P, Bret-Harte
805 MS, Comwell WK, Craine JM, Gurvich DE, others. 2013. New handbook for stand-
806 ardised measurement of plant functional traits worldwide. *Aust J Bot* **61**:167–234.

807 Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratorio S. 2016. Genome-
808 environment association study suggests local adaptation to climate at the regional scale
809 in Fagus sylvatica. *New Phytol* **210**:589–601. doi:10.1111/nph.13809

810 Poveromo JJ. 1999. Iron OresThe Making, Shaping, and Treating of Steel: Ironmaking
811 Volume. Pittsburg, PA: The AISE Steel Foundation. pp. 547–550.

812 Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Krämer U, Yant L. 2019. Convergent
813 evolution in Arabidopsis halleri and Arabidopsis arenosa on calamine metalliferous soils.
814 *Philos Trans R Soc B Biol Sci* **374**. doi:10.1098/rstb.2018.0243

815 Prober SM, Byrne M, McLean EH, Steane DA, Potts BM, Vaillancourt RE, Stock WD. 2015.
816 Climate-adjusted provenancing: a strategy for climate-resilient ecological restoration.
817 *Front Ecol Evol* **3**. doi:10.3389/fevo.2015.00065

818 Queiroz LP. 2015. Dioclea in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de
819 Janeiro. *http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB83091*.

820 Ramos SJ, Caldeira CF, Gastauer M, Costa DLP, Furtini Neto AE, de Souza FBM, Souza-
821 Filho PWM, Siqueira JO. 2019a. Native leguminous plants for mineland revegetation in
822 the eastern Amazon: seed characteristics and germination. *New For*. doi:10.1007/s11056-
823 019-09704-1

824 Ramos SJ, Gastauer M, Mitre SK, Caldeira CF, Silva JR, Furtini Neto AE, Oliveira G, Souza
825 Filho PWM, Siqueira JO. 2019b. Plant growth and nutrient use efficiency of two native
826 Fabaceae species for mineland revegetation in the eastern Amazon. *J For Res*.
827 doi:10.1007/s11676-019-01004-w

828 Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to
829 environmental association analysis in landscape genomics. *Mol Ecol* **24**:4348–4370.
830 doi:10.1111/mec.13322

831 Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the
832 spatial scale of evolution. *Trends Ecol Evol* **29**:165–176. doi:10.1016/j.tree.2014.01.002

833 Rossetto M, Bragg J, Kilian A, McPherson H, van der Merwe M, Wilson PD. 2019. Restore
834 and Renew: a genomics-era framework for species provenance delimitation. *Restor Ecol*
835 **27**:538–548. doi:10.1111/rec.12898

836 Russello MA, Waterhouse MD, Etter PD, Johnson EA. 2015. From promise to practice:
837 pairing non-invasive sampling with genomics in conservation. *PeerJ* **3**:e1106.
838 doi:10.7717/peerj.1106

839 Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G,
840 Dalén L, De Meester L, Ekblom R, Fawcett KD, Fior S, Hajibabaei M, Hill JA, Hoezel

AR, Höglund J, Jensen EL, Krause J, Kristensen TN, Krützen M, McKay JK, Norman AJ, Ogden R, Österling EM, Ouborg NJ, Piccolo J, Popović D, Primmer CR, Reed FA, Roumet M, Salmona J, Schenekar T, Schwartz MK, Segelbacher G, Senn H, Thaulow J, Valtonen M, Veale A, Vergeer P, Vijay N, Vilà C, Weissensteiner M, Wennerström L, Wheat CW, Zieliński P. 2015. Genomics and the challenging translation into conservation practice. *Trends Ecol Evol* **30**:78–87. doi:10.1016/j.tree.2014.11.009

Shryock DF, Havrilla CA, DeFalco LA, Esque TC, Custer NA, Wood TE. 2017. Landscape genetic approaches to guide native plant restoration in the Mojave Desert. *Ecol Appl* **27**:429–445. doi:10.1002/eap.1447

Shryock DF, Havrilla CA, DeFalco LA, Esque TC, Custer NA, Wood TE. 2015. Landscape genomics of Sphaeralcea ambigua in the Mojave Desert: a multivariate, spatially-explicit approach to guide ecological restoration. *Conserv Genet* **16**:1303–1317. doi:10.1007/s10592-015-0741-1

Silva AR, Resende-Moreira LC, Carvalho CS, Lanes ECM, Ortiz-Vera MP, Viana PL, Jaffé R. 2020. Range-wide neutral and adaptive genetic structure of an endemic herb from Amazonian Savannas. *AoB Plants* **12**. doi:10.1093/aobpla/plaa003

Silva JR, Gastauer M, Ramos SJ, Mitre SK, Furtini Neto AE, Siqueira JO, Caldeira CF. 2018. Initial growth of Fabaceae species: Combined effects of topsoil and fertilizer application for mineland revegetation. *Flora Morphol Distrib Funct Ecol Plants* **246–247**:109–117. doi:10.1016/j.flora.2018.08.001

Skirycz A, Castilho A, Chaparro C, Carvalho N, Tzotzos G, Siqueira JO. 2014. Canga biodiversity, a matter of mining. *Front Plant Sci* **5**:653. doi:10.3389/fpls.2014.00653

Souza-Filho PWM, Giannini TC, Jaffé R, Giulietti AM, Santos DC, Nascimento Jr. WR, Guimarães JTF, Costa MF, Imperatriz- Fonseca VL, Siqueira JO. 2019. Mapping and quantification of ferruginous outcrop savannas in the Brazilian Amazon: A challenge for biodiversity conservation. *PLoS One* **14**:e0211095.

Steane DA, Potts BM, McLean E, Prober SM, Stock WD, Vaillancourt RE, Byrne M. 2014. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Mol Ecol* **23**:2500–2513. doi:10.1111/mec.12751

Supple MA, Bragg JG, Broadhurst LM, Nicotra AB, Byrne M, Andrew RL, Widdup A, Aitken NC, Borevitz JO. 2018. Landscape genomic prediction for restoration of a Eucalyptus foundation species under climate change. *Elife* **7**:1–22. doi:10.7554/eLife.31835

Talbot B, Chen T-W, Zimmerman S, Joost S, Eckert AJ, Crow TM, Semizer-Cuming D, Seshadri C, Manel S. 2016. Combining Genotype, Phenotype, and Environment to Infer Potential Candidate Genes. *J Hered* **108**:esw077. doi:10.1093/jhered/esw077

876 Vangestel C, Eckert AJ, Wegrzyn JL, St. Clair JB, Neale DB. 2018. Linking phenotype,
877 genotype and environment to unravel genetic components underlying cold hardiness in
878 coastal Douglas-fir (Pseudotsuga menziesii var. menziesii). *Tree Genet Genomes* **14**:10.
879 doi:10.1007/s11295-017-1225-x

880 Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, Byrne M, Coates DJ,
881 Eldridge MDB, Sunnucks P, Breed MF, James EA, Hoffmann AA. 2011. Assessing the
882 benefits and risks of translocations in changing environments: A genetic perspective.
883 *Evol Appl* **4**:709–725. doi:10.1111/j.1752-4571.2011.00192.x

884 Whiting SN, Reeves RD, Richards D, Johnson MS, Cooke JA, Malaisse F, Paton A, Smith
885 JAC, Angle JS, Chaney RL, Ginocchio R, Jaffre T, Johns R, McIntyre T, Purvis OW, Salt
886 DE, Schat H, Zhao FJ, Baker AJM. 2004. Research Priorities for Conservation of
887 Metallophyte Biodiversity and their Potential for Restoration and Site Remediation.
888 *Restor Ecol* **12**:106–116. doi:10.1111/j.1061-2971.2004.00367.x

889 Williams A V., Nevill PG, Krauss SL. 2014. Next generation restoration genetics:
890 Applications and opportunities. *Trends Plant Sci* **19**:529–537.
891 doi:10.1016/j.tplants.2014.03.011

892 Xuereb A, Kimber CM, Curtis JMR, Bernatchez L, Fortin MJ. 2018. Putatively adaptive
893 genetic variation in the giant California sea cucumber (Parastichopus californicus) as
894 revealed by environmental association analysis of restriction-site associated DNA
895 sequencing data. *Mol Ecol* **27**:5035–5048. doi:10.1111/mec.14942

896 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath
897 AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. 2010. Common SNPs
898 explain a large proportion of the heritability for human height. *Nat Genet* **42**:565–569.
899 doi:10.1038/ng.608

900 Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA,
901 Smets P, Wang T, Gray LK, Liepe KJ, Hamann A, Holliday JA, Whitlock MC, Rieseberg
902 LH, Aitken SN. 2016. Convergent local adaptation to climate in distantly related
903 conifers. *Science (80-)* **353**:1431–1433. doi:10.1126/science.aaf7812

904 Zappi DC, Moro MF, Walker B, Meagher T, Viana PL, Mota NFO, Watanabe MTC, Nic
905 Lughadha E. 2019. Plotting a future for Amazonian canga vegetation in a campo rupestre
906 context. *PLoS One* **14**:e0219753. doi:10.1371/journal.pone.0219753
907
908
909

Data Availability Statement

Geographic coordinates, genotypes in Variant Call Format, and sequences in FASTA format for both species are available in Figshare: https://doi.org/10.6084/m9.figshare.12185235.v1. All the mentioned R scripts have been deposited in GitHub and their url addresses provided in the text.

Author contributions

RJ conceived, designed and coordinated the project. CSC, MG and RJ coordinated the field work and sampling. CSC, SM and MG performed laboratory work. CSC, LCRM, LT, BF, MG and RJ performed the data analysis. The first draft of the paper was written by CSC and RJ and all authors contributed to discussing the results and editing the paper.

Tables

Table 1: Genetic diversity measures for *Mimosa acutistipula* var. *ferrea* and *Dioclea apurensis* within genetic clusters (highlands). Sample sizes (*N*) are followed by mean expected heterozygosity (H_E), mean inbreeding coefficient (*F*), nucleotide diversity (π) and effective population size (N_e). Values represent 95% confidence intervals.

Species	Highland	*N*	H_E	*F*	π	N_e
M. acutistipula	Serra Sul	80	[0.28/0.28]	[0.12/0.15]	[0.26/0.26]	[1601.0/1778.0]
	Serra Norte	61	[0.28/0.28]	[0.10/0.14]	[0.23/0.24]	[554.7/581.2]
	Serra da Bocaina	36	[0.28/0.28]	[0.12/0.17]	[0.25/0.26]	[1008.0/1243.5]
	Total	177	[0.28/0.28]	[0.10/0.17]	[0.23/0.26]	[554.7/1778.0]
D. apurensis	Serra Sul	81	[0.25/0.25]	[0.10/0.14]	[0.22/0.23]	[1070.1/1278.5]
	Serra Norte	45	[0.29/0.29]	[0.14/0.20]	[0.24/0.25]	[193.2/204.6]
	Serra da Bocaina	37	[0.27/0.27]	[0.22/0.31]	[0.21/0.22]	[193.7/212.2]
	Total	163	[0.25/0.29]	[0.10/0.31]	[0.21/0.25]	[193.2/1278.5]

Figures

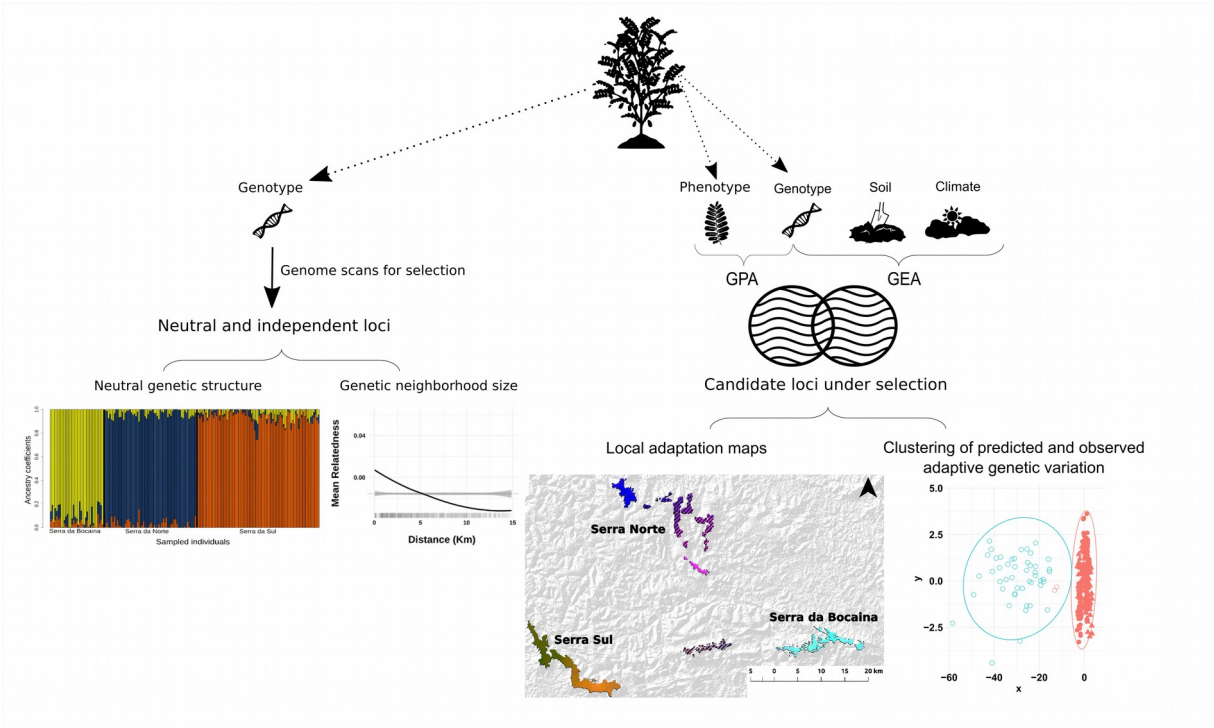
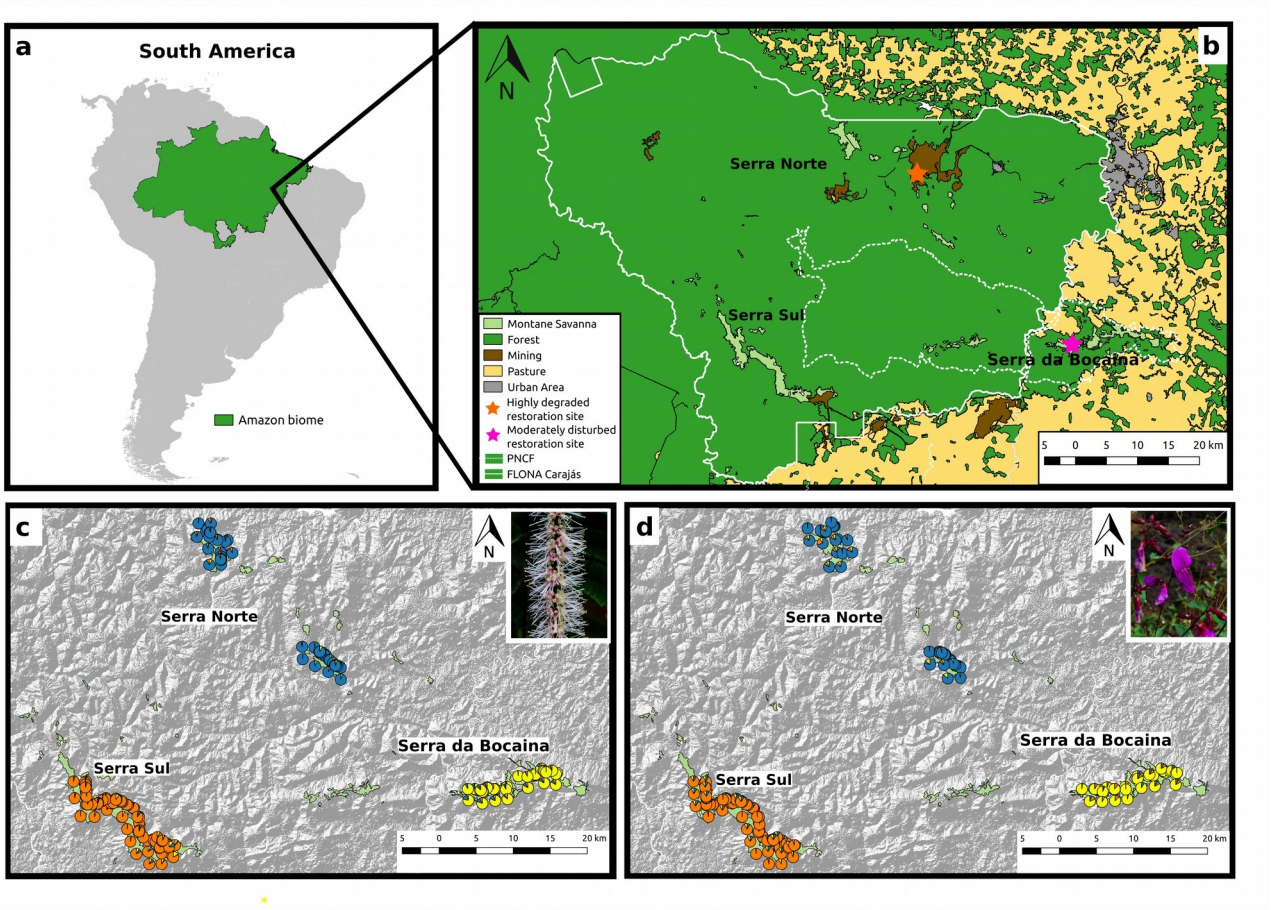
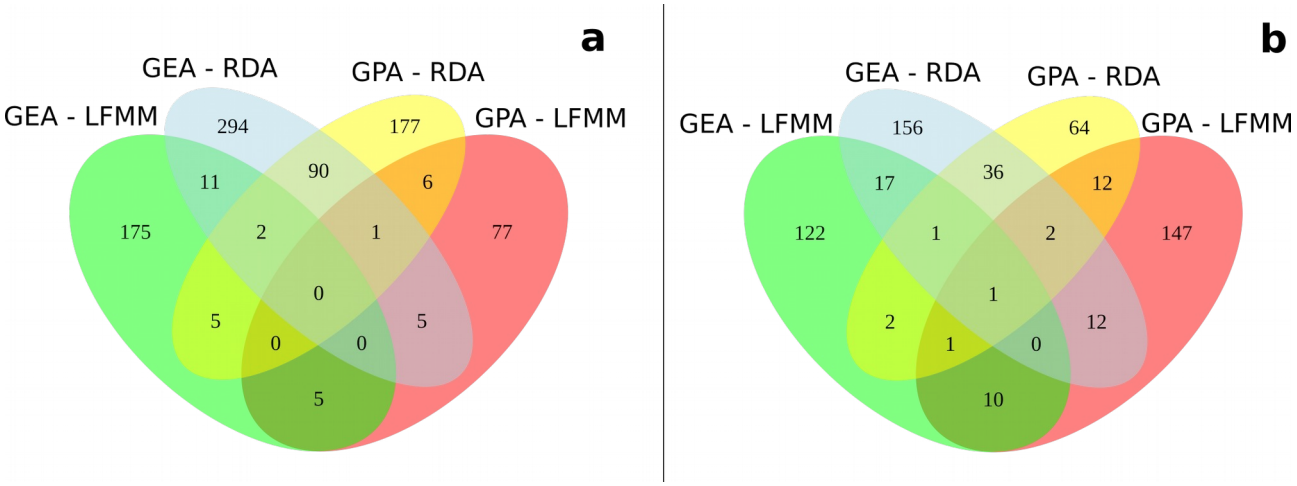


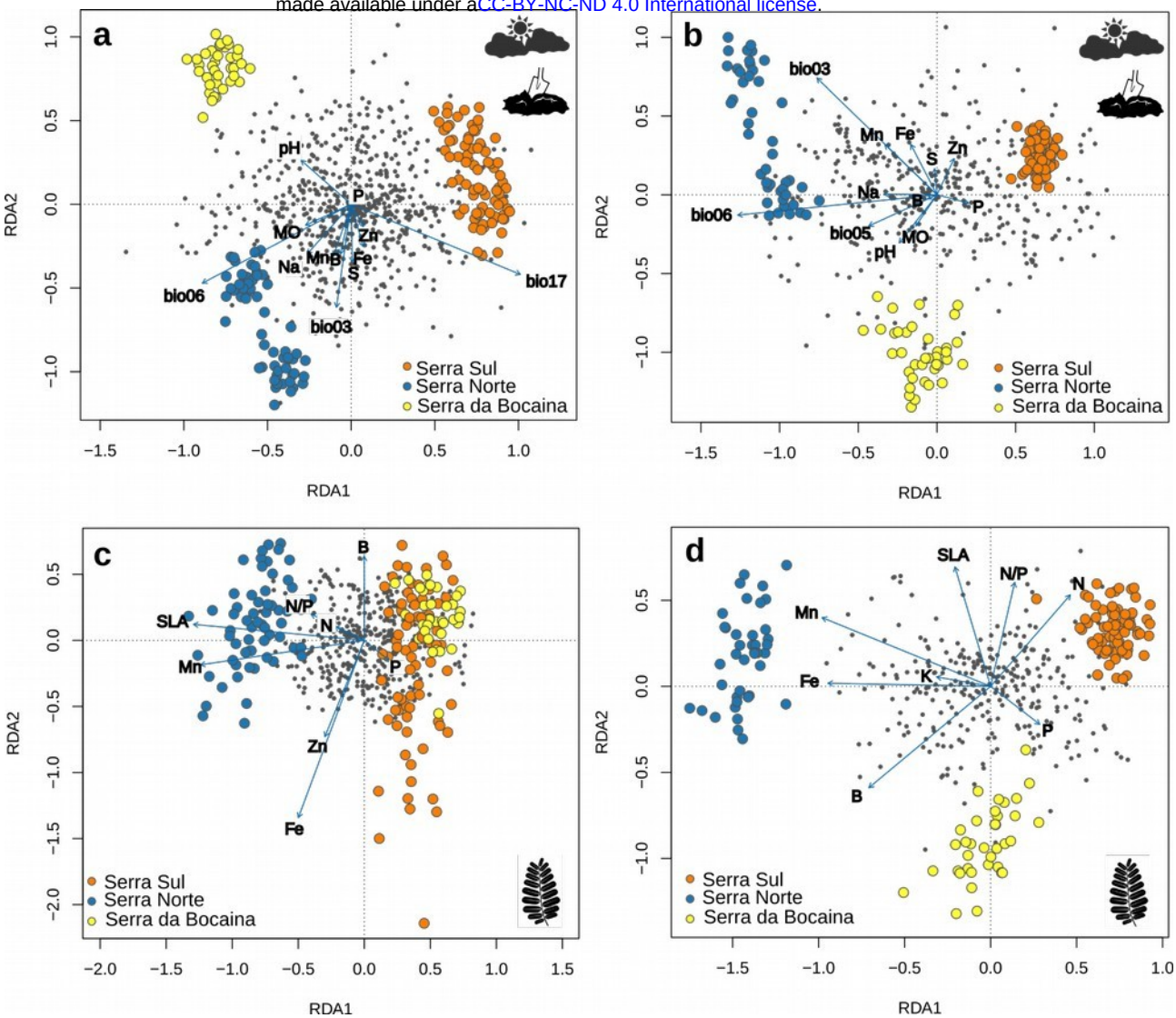
Figure 1: Diagram summarizing our methodological approach. We used a subset of neutral and independent loci to assess broad and fine-scale genetic structure and determine the genetic neighborhood size. Subsequently, we combined genotype, phenotype, and environmental data to identify loci under selection and then employed all candidate loci to map patterns of adaptive genetic variability and predict adaptive genotypes for restoration sites. The graphs show results for *Mimosa acutistipula.*



938 **Figure 2:** Maps of our study region depicting restoration sites and neutral genetic structure
939 for *Mimosa acutistipula* var. *ferrea* and *Dioclea apurensis*. **a**: Location of the Carajás Mineral
940 Province within the Amazon biome. **b:** Location of the major Canga highlands (Montane
941 Savanna) where samples were collected. White continuous lines represent the Carajás
942 National Forest (FLONA Carajás), white dashed lines the Campos Ferruginosos National Park
943 (PNCF), and stars depict restoration sites. Ancestry coefficients for all samples from *M.*
944 *acutistipula* var. *ferrea* (**c**) and *D. apurensis* (**d**), determined using the *snmf* function from the
945 LEA package.



946 **Figure 3:** Venn diagram showing the intersection of contigs (RAD-tags) containing candidate
947 SNPs for *Mimosa acutistipula* var. *ferrea* (**a**) and *Dioclea apurensis* (**b**). Genotype-
948 environment associations (GEA) and genotype-phenotype association (GPA) were assessed
949 using Redundant Analysis (RDA) and Latent Factor Mixed Model (LFMM). The number of
950 detections by method for each species are presented in Tables S2 and S3.



951 **Figure 4:** Redundancy analysis (RDA) using only the candidate loci identified in genotype-environment
952 (upper panels) and genotype-phenotype association analyses (lower panels) in *Mimosa acutistipula* var.
953 *ferrea* (**a**, **c**) and *Dioclea apurensis* (**b**, **d**). The plots show the first and second constrained axes from
954 RDA, with SNPs represented as gray filled circles, environment and phenotype variables as blue arrows,
955 and individuals from different Canga highlands in color circles (bio03: isothermality; bio05: maximum
956 temperature of warmest month; bio06: minimum temperature of coldest month; bio17: precipitation of
957 driest quarter; MO: organic material; and SLA: specific leaf area).

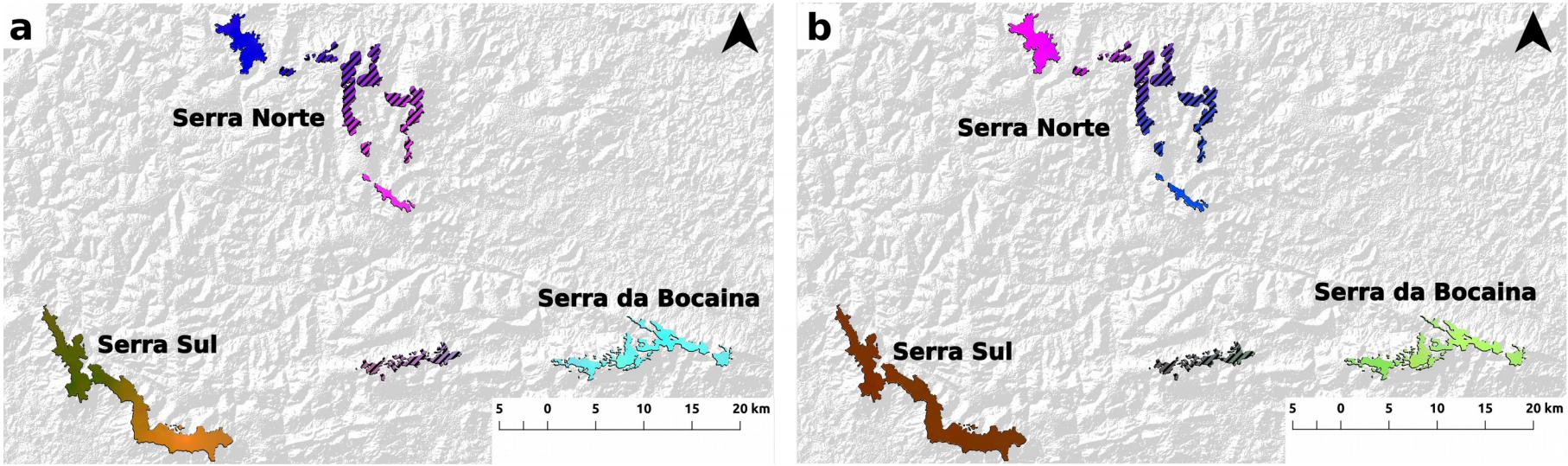


Figure 5: Spatial distribution of adaptive genetic variation in *Mimosa acutistipula* var. *ferrea* (**a**) and *Dioclea apurensis* (**b**). The maps represent an RGB composite made using interpolated principal components from a sPCA, ran on the combined candidate loci found in GEA and GPA. Regions with similar colors within each panel represent analogous genetic composition and areas with diagonal lines were not sampled (i.e., adaptive genetic composition was extrapolated from neighboring areas).

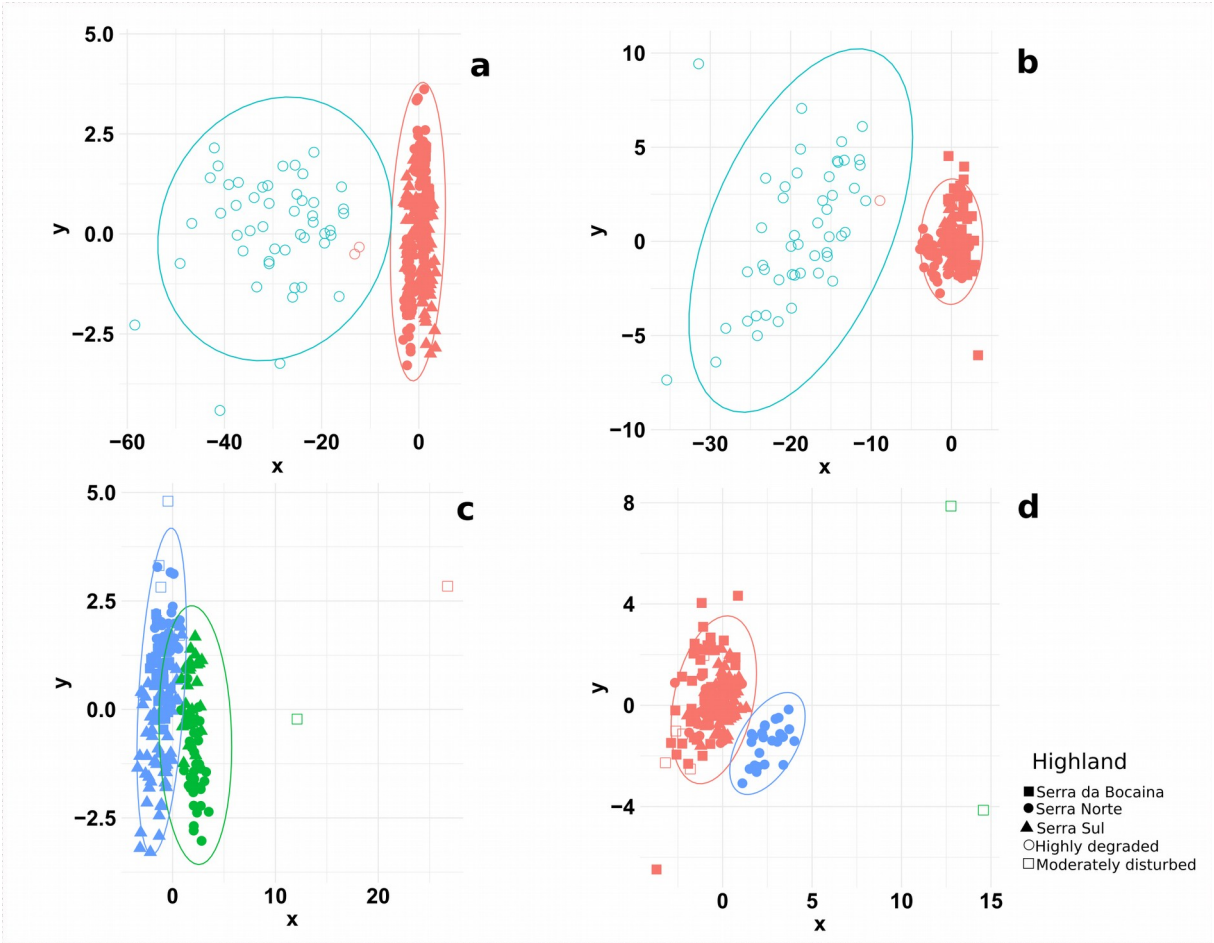


Figure 6: Clustering of predicted and observed adaptive genetic variation in *Mimosa acutistipula* var. *ferrea* (a, c) and *Dioclea apurensis* (b, d). A *k*-means clustering approach was employed to examine the similarity between observed (filled symbols) and predicted genotypes (open symbols) associated with the environmental conditions of highly degraded (upper panels) and moderately disturbed (lower panels) sites. Colors indicate different clusters and symbols the locations where samples were collected.